

How can I calculate the cumulative sum by group using Pandas?

Authored by
stats writer

June 30, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I calculate the cumulative sum by group using Pandas?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=160685>

Pandas is a widely used Python library for data manipulation and analysis. It offers powerful tools for grouping and aggregating data, allowing for efficient data exploration. One useful function in Pandas is the ability to calculate the cumulative sum by group. This means that the sum of a particular variable will be calculated for each group within a dataset, and the results will be displayed in a cumulative manner. This can be achieved by using the 'groupby' function in Pandas, which allows for grouping data by a certain variable, followed by the 'cumsum' function, which calculates the cumulative sum for each group. This feature in Pandas is beneficial for analyzing data in a more organized and insightful manner, providing valuable information for decision making and further analysis.

Pandas: Calculate Cumulative Sum by Group

You can use the following syntax to calculate a cumulative sum by group in pandas:

```
df = df.groupby().cumsum()
```

This particular formula calculates the cumulative sum of col2, grouped by col1, and displays the results in a new column titled cumsum_col.

The following example shows how to use this syntax in practice.

Example: Calculate Cumulative Sum by Group in Pandas

Suppose we have the following pandas DataFrame that contains information about sales for various stores:

```
import pandas as pd
```

```
#create DataFrame
```

```
df = pd.DataFrame({'store': ,  
'sales': })
```

```
#view DataFrame
```

```
print(df)
```

```
store sales
```

```
0 A 4
```

```
1 A 7
```

```
2 A 10
```

```
3 A 5
```

```
4 A 8
```

```
5 B 9
```

```
6 B 12
```

```
7 B 15
```

```
8 B 10
```

```
9 B 8
```

We can use the following syntax to calculate the cumulative sum of sales for each store:

```
#add column that shows cumulative sum of sales by  
store
```

```
df = df.groupby().cumsum()
```

```
#view updated DataFrame  
print(df)
```

```
store sales cumsum_sales
```

```
0 A 4 4  
1 A 7 11  
2 A 10 21  
3 A 5 26  
4 A 8 34  
5 B 9 9  
6 B 12 21  
7 B 15 36  
8 B 10 46  
9 B 8 54
```

The `cumsum_sales` column shows the cumulative sales, grouped by each store.

Note: You can find the complete documentation for the `cumsum` function in pandas .

Additional Resources

The following tutorials explain how to perform other common tasks in pandas: